



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

Zucon, Guido (2009) An analogy between the double slit experiment and document ranking. In *3rd IRSG Symposium : Future Directions in Information Access*, 1 September 2009, Padua, Italy.

This file was downloaded from: <http://eprints.qut.edu.au/69279/>

© Copyright 2009 [please consult the author]

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

An Analogy between the Double Slit Experiment and Document Ranking

Guido Zuccon
Dept. of Computing Science
University of Glasgow (Scotland)

Abstract

In the last years several works have investigated a formal model for Information Retrieval (IR) based on the mathematical formalism underlying quantum theory [1, 2, 3, 4, 5]. These works have mainly exploited geometric and logical–algebraic features of the quantum formalism, for example entanglement, superposition of states, collapse into basis states, lattice relationships. In this poster I present an analogy between a typical IR scenario and the double slit experiment. This experiment exhibits the presence of interference phenomena between events in a quantum system, causing the Kolmogorovian law of total probability to fail. The analogy allows to put forward the routes for the application of quantum probability theory in IR. However, several questions need still to be addressed; they will be the subject of my PhD research.

Keywords: document ranking, interdependent document relevance, quantum probability theory, interference, formal models

1. INTRODUCTION

Recently there has been an increasing interest on formal IR models inspired by the mathematical formalism of quantum theory. Some classical IR models, such as the vector space model, the probabilistic model, the logical model [3], and the logical imaging technique [5] have been expressed into the Hilbert space framework proper of quantum theory. In [4], the concepts of non-relevancy and negation of a term have been translated in orthogonality between subspaces of a Hilbert space. In [1], the role of context for word association is modeled by means of direct entanglement between words in high dimensional semantic spaces. Logic relationships between terms and the context surrounding them has been introduced in [2] by means of transformations on the text, the selective erasers, inspired by quantum measurements.

However, no previous work has been focusing on the nature of probabilities in IR. In particular, the use of quantum probability theory as feasible modeling tool for IR has never been explored. In the following, I introduce an analogy between the double slit experiment and a user examining a ranking of documents that have been retrieved by an IR system in response to his/her information need. The double slit experiment exhibits the arising of the interference phenomena; the presence of interference causes the violation of the Kolmogorovian law of total probability.

2. THE ANALOGY

The double slit experiment consists of shooting a physical particle towards a screen with two slits, named A and B (Fig. 1(a)). Once the particle passes through one of the slits, it hits a detector panel, positioned behind the screen, in a particular location x with probability $p_{AB}(x)$. The (complex) probability amplitude associated to the events of passing through A (alternatively, B) when slit B (A) is closed and being detected at x is indicated by $\phi_A(x)$ ($\phi_B(x)$). Amplitude probabilities are linked to probabilities by the following equations: $p_A(x) = |\phi_A(x)|^2$, $p_B(x) = |\phi_B(x)|^2$. Intuitively, we would expect that the probability of the particle being detected at x when both slits are open is the sum of the probability of passing through A and being detected at x , $p_A(x)$, and the probability of passing through B and hit the detector panel in x , $p_B(x)$. However, experimentally it has been noted that $p_{AB}(x) \neq p_A(x) + p_B(x)$. Instead, the probability distribution

obtained measuring $p_{AB}(x)$ across the detection panel presents an interference pattern akin to waves that would pass through both slits and hit the detector panel. In particular, the (complex) probability amplitude of a particle being measured at position x after passing through either slit A or B (indicated as $\phi_{AB}(x)$) is the sum of the probability *amplitude* associated to the event of opening just slit A plus the counterpart event of having open just slit B : $\phi_{AB}(x) = \phi_A(x) + \phi_B(x)$. In terms of probabilities,

$$p_{AB}(x) = |\phi_{AB}(x)|^2 = |\phi_A(x)|^2 + |\phi_B(x)|^2 + (\phi_A(x)^* \phi_B(x) + \phi_A(x) \phi_B(x)^*) = p_A(x) + p_B(x) + I_{AB}(x) \quad (1)$$

where term $I_{AB}(x)$ represents the quantum *interference* term, which is modulated by the phase difference between the amplitudes $\phi_A(x)$ and $\phi_B(x)$. In fact, by expanding $I_{AB}(x)$ and letting θ_{AB} being the phase difference between the probability amplitudes $\phi_A(x)$ and $\phi_B(x)$, we obtain

$$I_{AB}(x) = \phi_A(x)^* \phi_B(x) + \phi_A(x) \phi_B(x)^* = 2 |\phi_A(x)| |\phi_B(x)| \cos \theta_{AB} \quad (2)$$

The analogy between the double slit experiment and the IR situation follows. The particle is associated with the user and his information need, while each slit represents a document (Fig. 1(b)). The event of passing through a slit is seen as the action of examining the ranking of documents, e.g. read the associated snippets or the documents themselves. Measuring at x stands for assessing the satisfaction of the user given the list of documents, or more concretely the decision of the user to stop his search (event x , the user is fully satisfied) or continue searching (\bar{x} , he is not completely satisfied by the documents). In these settings, being detected with probability $p_{AB}(x)$ at position x on the panel means choosing to stop the search with probability $p_{AB}(x)$ after being presented with documents A and B . Analogously to the double slit experiment, I propose that in the IR scenario the probability of the user being satisfied by the ranking of documents A and B is given by the sum of the probability of the single events (satisfied by document A , satisfied by document B) and the additional probability associated to the interference term. Thus, the satisfaction of the user does not depend just upon the relevance/satisfaction provided by each document independently (as it is commonly assumed in IR). Conversely, it is affected by the interference between the relevance/satisfaction of the entire document ranking. This suggests that a model of document ranking based on quantum probabilities might exploit interdependent document ranking.

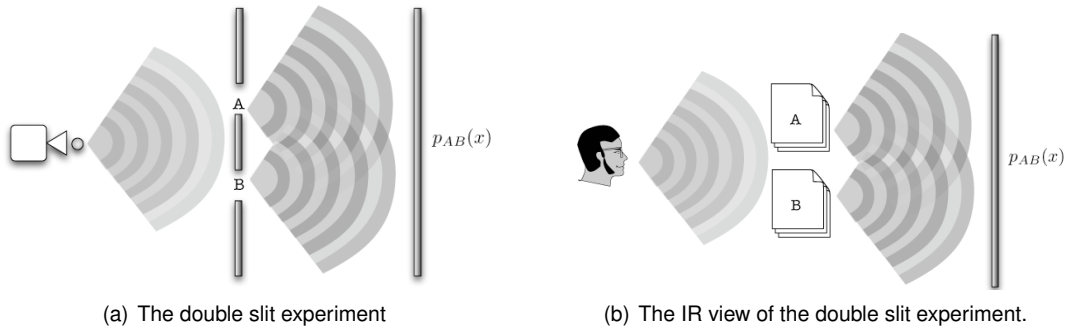


FIGURE 1: Schematic representation of the analogy between the double slit experiment and the IR document ranking problem.

3. CONCLUSIONS

In this poster I have presented an analogy between the double slit experiment and a typical IR scenario. The analogy suggests that when calculating the probability of a document ranking being relevant to an user's information need the interference between the relevancy of the documents themselves should be accounted for. At this stage, several questions need to be investigated and they will be part of my PhD research. Among those, the most urgent are: What are the implications for IR? What do complex probabilities mean in IR? What does the interference term represent in IR? What is the behaviour of the interference term? How the interference term can be computed in IR?

4. ACKNOWLEDGEMENTS

This research has been sponsored in part by EPSRC project Renaissance (EP/F014384/1) and Royal Society International Joint Project JP080734.

REFERENCES

- [1] P. D. Bruza, K. Kitto, D. L. Nelson, and C. L. McEvoy. Entangling words and meaning. In *Second Quantum Interaction Symposium*, 2008.
- [2] A.F. Huertas-Rosero, L.A. Azzopardi, and C.J. van Rijsbergen. Eraser lattices and semantic contents: An exploration of the semantic contents in order relations between erasers. In *LNAI Series: LNCS, volume 5494*, pages 266–275. Springer Verlag, 2009.
- [3] C. J. van Rijsbergen. *The Geometry of Information Retrieval*. Cambridge Univ. Press, NY, USA, 2004.
- [4] D. Widdows. *Geometry and Meaning*. CLSI Lec. Not. CSLI, 2004.
- [5] G. Zuccon, L. A. Azzopardi, and C. J. van Rijsbergen. A formalization of logical imaging for information retrieval using quantum theory. In *IEEE Proc. 5th Int. Workshop TIR*. IEEE, 2008.